SEQUENCE LISTING

(1) GENERAL INFORMATION:

    (i) APPLICANT: Hanna, Michael C.
                  Kirkness, Ewen F.

    (ii) TITLE OF INVENTION: $GABA_A$ Receptor Epsilon Subunits

    (iii) NUMBER OF SEQUENCES: 46

    (iv) CORRESPONDENCE ADDRESS:
        (A) ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
        (B) STREET: 1100 New York Avenue, NW, Suite 600
        (C) CITY: Washington
        (D) STATE: DC
        (E) COUNTRY: USA
        (F) ZIP: 20005-3934

    (v) COMPUTER READABLE FORM:
        (A) MEDIUM TYPE: Floppy disk
        (B) COMPUTER: IBM PC compatible
        (C) OPERATING SYSTEM: PC-DOS/MS-DOS
        (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

    (vi) CURRENT APPLICATION DATA:
        (A) APPLICATION NUMBER: To be assigned
        (B) FILING DATE: Herewith
        (C) CLASSIFICATION:

    (vii) PRIOR APPLICATION DATA:
        (A) APPLICATION NUMBER: US 08/888,012
        (B) FILING DATE: 03-JUL-1997
        (C) CLASSIFICATION:

    (viii) ATTORNEY/AGENT INFORMATION:
        (A) NAME: Steffe, Eric K.
        (B) REGISTRATION NUMBER: 36,688
        (C) REFERENCE/DOCKET NUMBER: 1488.0950001/EKS/SGW

    (ix) TELECOMMUNICATION INFORMATION:
        (A) TELEPHONE: (202) 371-2600
        (B) TELEFAX: (202) 371-2540

(2) INFORMATION FOR SEQ ID NO:1:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 6146 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (ix) FEATURE:
        (A) NAME/KEY: CDS
        (B) LOCATION: 3872..4597

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
GTCTTATAAT TGCTAAGCAC TTACAACTGT TTGCAGAGGA AACTGAGACT TTGTAACTAT      60

GTCTCAGTCT CATCTGCAAA GAAGTAAGTG CTTTGCCAAG CTCCTTGAGA GGTTAGGTAA     120

GTAGATAAAG TTCTGCTGCT GTCGGAATGT GCAGCTGGCT TTTTCATGCA GACCCTTCAG     180

TTTCGAGGTT ACAACTCTGA CCTCTTTGGA TGACTTTGGG GAATGGAGCT CGTGTGAGTT     240

CTCCATACCC AGAACCAATC CAGTCTGGTT GAATGGGAAG CAAAGTCCAT TGTAGTGGGA     300

GGTGGAGGCT AGAGTTCTAA TGTCAGCTAG TTTAAGGCTG GGAAAGTCTG GAGGAAGTTA     360

CAGCAGCTAC ACTGGCTGCT GCATTGACAT TTATCTTAAA GGAACAAGTC TGAAAAGCAC     420

AGATTCTTAT CAAAGGCTTC ATGGTGGATT CCACATAGAC ATAGTGGCCA CTGGTTTTCT     480

GACCTTTTCT CTGACAAAGA CTAAAGGGGA AGGTCCTGGG TATCTTACAC TTCAGCTCCC     540·

AATTAGATGT GAGCACCTTC ACTTATGTTC CTAGGTGACC TGAATGAGGA GCCAAGGGAC     600

CTCCCCAGGG TAGCTCCCAG AGCAACCCTG GAAACACTCT TCACACATCC TGACCAAGTT     660

CAGGGCAGTG AAGGCACTGC CCTCATCGTT TCCAGAATGT GGATGGAGCC AGTCACCCAA     720

CCAGCCATTT GTCGTGAGAG GCATCTTGTT CTGCTACCAT GTGACTAGGC AGAAAATCTG     780

CTTTTGTTTC ATTTATTGAG TCAGTCTCTG GATGAGGGAA AGCTCATGCT CATGTGGCTA     840

GAGCTTTGCT TGCACAGTAT TAGGCAGGGG CAGAGGGCTG GGCTACCTTA AAAATACTTG     900

CCCTTTTTCT TGGGGACTCT GGGGAAGCGG TTTTACTACC TTTGACTTGG GAGCCTTGCT     960

CTTCTGCCAG CTAACCATGG GCCTGCCTCT TGGTTTTCTG CACCTCAGCT TTTCCCGGAT    1020

AGGTGGGGAC CCATCATCAA AAGTGACAGA GAAGATAAGG CCCAGGGGCT TTCAAGTCAC    1080

TAGTGGTTCC GTTTAGTAGA TGATTGTGCA TTGTTTCAAA ATGGTGCCCT AGTGACTACA    1140

AAGCCCCAGA GCCAGCATCA TCATCAAAGC AATGACAGTA GGTAAGCACC AGACCTCCTT    1200

GGGAGTGAGG AGGATTCTTG AGGAGAAAAG AGGTCTTCTT TCTCCTCTGC TGGAGACTAG    1260

TTGATCTGGA GACGTGGTTC CTTCAATGTC AGAGTTATCT TTGGGACTGG TCTCAAACTC    1320

TTCCAGTTGG GCCCTGGGGC AGGTCTCTCC ATCTGGAGCA TACTTACGTG CTCGGCGATT    1380

AAGGGTTCAG AATGCAGTGG TAGCCTGCTA CTCTGGCCAT CTTGGACCTT GATCCAGAGA    1440

ATCTCTGCTT CAGGAGCTTC TAAGAGAGTC CAGCCCTGCC TCCAGAGAGA GGCTTGCCCT    1500

TCACTGATGG CTGTGGAGCC TCTGATGGAA TATTATTGCT GGTCAGGAAT TCACTGTCTT    1560

ACAAGGAGGT TTCCTTCTTC TCTAGACAGT TCTGTTCATC AAAAAACTCT CCCTGTTCTT    1620

CTGAAATTGG AGTCTCTGGA AGTTCCACAC ATTAAGCTTA GTTCTTTTTC CTTGGAACTG    1680

TCCAGGTTAC ATTAGTCCAG CCACTGTTTC ACAGGACCGA GATTAAACGA TCAACATCAT    1740

CATTCCCGGC ATGGATCATA GTCTGTTGTA GTCTACATAG CCCTAGTTTA TTTTTCTTCC    1800
```

```
CTTATTCTTC AAAGCTTTGG GTCCATTCAT TCTTCTAGTC CCAGTCCTCT GGACATGGTC    1860

TATTTAATTG TGTCCCTCTG ACACTGCAGT GACCAACCAT GATCTGGTCA AAGAGGATAA    1920

GAGTTTGAGC AGAAAACCAT CTTTAGCATA TATTTTTTTG CTTTGGTTCA TCAGCCCCAG    1980

ATATATTGTT TTCCTTACCC GTGCTTCTCT CACTCCTCAA GAAGAAGAAA GTGTGTGTTA    2040

GCATCTTTCT CTTGTCCTTC AAGACAAATT GGCATCTCTT GACGAGCGGA GAAGGTTCTT    2100

TTTTGGCCAG AATAAATAAA ATTAAAATAG AATCATCCAA CAGAATAATA AATCTTCGTG    2160

CAACAAGAAT ATATTATATA AACCCAGCAA TTTTGCAGGG CCTGGGTATA ACTAATTAGA    2220

AGTGTCTTAA ATTGCAGTCA AGATCCCACG GCAAGAGGAC TTTTGATAAA TACATTCTGG    2280

CCAGTAGGCA AGTGCGAGGG TGGTCCGTGC AGCAGCTCTG GAGGAGTTCT ATCCCAAAGC    2340

TATACTCAAC ACACAGGTTT CCCACTGACA ACAGGTCGCT CCCTTGCCTT CTTCCAGAAG    2400

AATCTGAGAA GCTTTGCTCC TTGAGTTTCA GTGCTGCCAA GGTGAGTACG AAAGGCTGCT    2460

CTTCTCATTC AGCTCCAGCC CACCCAGACC TGCTGGGCAG TTGATCCACT TTCCAAAATA    2520

GGAGGACACA CGGACAGGTT AGTGTTCTGG TCTGCTTTAC AAAGCTGTTG CCTGACAGGA    2580

GCAAGAGTTG CTGAGTGTCT GCTGGGTTCC AGGCTGTTCT GAGCTTGGAT GGGCAGGGGC    2640

TAAGCCACAG GGCCTGCATG AGCCCTGCCT TGAAGGGACT TAAAAGACGA CCTAATTATA    2700

GGCCTAGGAA TTTTACAGTA TTGCAACTGC AATGTGATGC TGAAAGTGGA AAATGATGTC    2760

CTGGGCTCAG AGAAAAGCCC ACACCAGCCT GGGAGTCATG ATAGCAGCAG AGTGCTTGGG    2820

GAGGGTGTGT CAGAGCATAA AGCAGCATGA ATGCTACAAA GAAGATGCC AACTAGAGAT    2880

ATAGGTTGTC ATCAGGTCCC GGAGGAGCCA TGACCGTCTA GCTGAGAGCC ATGACCAAGG    2940

ACACAATGTC CAAGTGACTG TGAGGACCTC AGTCTGCCCT GTGGATGTGT ATGCCACAGA    3000

CCTGACTTCT GGAGGGCTGA CTGAAATGTT CATTTTAAGC TTTTTCTTCT CTTTCCCTGA    3060

AACACTCAGT TTGGGTTAGG GGTCATAGAC TAAGACCAAA GAGTCCAGGG TTAGAATCTT    3120

GGTGTAAAAT TGCAGGCCAT CTCAGGAAAT CTGTGAGCAG ATGGGATTGG CTTTGGGTAA    3180

GGTGCGTGTG GAAAATGTCA GTGGGAGCCG GGTCATGGTG GGCCTTTAGC ATCAGATTCC    3240

AGAGTGCAGA TAGTCTGTAT AGCTCATGTG AAACAGGGAG CCACCAAAAC TTTGGGGAGC    3300

AGGCTAGTGC CGGTTTTGAC CACCTGTGGA GCAGTGCTCA CTCACGAAGG CATTTTGCCA    3360

TCACATGAAT GTGCAGAAAG GAGGCCAAAA GCATTCTGTG CTTCTCCACC ACAGCACAGA    3420

CTTCCCTAGT CTCATTTGCT GAGAGTAGAC ATTCTGAGGG CCAGCAGTGC AGGTGTGATG    3480

TGCCTCAGAG GGTATGAAGC CCTTAGTCAG CCATCTGGAT ATCAGCTGCG TGGGCATGAT    3540

ATCTAGAAGG CTAATTGATT TTTTCACTTT CACCTGACTC TCTTGCCAAC CTGCAGAGAC    3600

AGACATTGGG TGTAGGACAG TGAACTGAGA AGGAAGCTAT TAAGATTCTG GCCTTGGCTT    3660
```

```
AGCTCTCAAC TGGCCATTGG TCTTGCAGTA AGTCTTTTTT CTGGGCTTCT TCTGGTCCTA        3720

TTTGTATGTA TTGCATTGTC ACATCATGCC TCTATCCTAG GGAATACTGT GAGCTGAAAA        3780

ATGAGACCCT TACTGTTCAC GTCCTGCTAA GGGGGACCGT CGTGTCAGCA CTGTAATGCA        3840

GTGATGTTTT TTGTGTCTTT CAGGTGACTT C ATG GTC ATG ACG ATT TTC TTC          3892
                                   Met Val Met Thr Ile Phe Phe
                                    1                   5
```

```
AAT GTG AGC AGG CGG TTT GGC TAT GTT GCC TTT CAA AAC TAT GTC CCT        3940
Asn Val Ser Arg Arg Phe Gly Tyr Val Ala Phe Gln Asn Tyr Val Pro
         10              15                  20
```

```
TCT TCC GTG ACC ACG ATG CTC TCC TGG GTT TCC TTT TGG ATC AAG ACA        3988
Ser Ser Val Thr Thr Met Leu Ser Trp Val Ser Phe Trp Ile Lys Thr
     25              30                  35
```

```
GAG TCT GCT CCA GCC CGG ACC TCT CTA GGG ATC ACC TCT GTT CTG ACC        4036
Glu Ser Ala Pro Ala Arg Thr Ser Leu Gly Ile Thr Ser Val Leu Thr
 40                  45              50                      55
```

```
ATG ACC ACG TTG GGC ACC TTT TCT CGT AAG AAT TTC CCG CGT GTC TCC        4084
Met Thr Thr Leu Gly Thr Phe Ser Arg Lys Asn Phe Pro Arg Val Ser
                 60              65                  70
```

```
TAT ATC ACA GCC TTG GAT TTC TAT ATC GCC ATC TGC TTC GTC TTC TGC        4132
Tyr Ile Thr Ala Leu Asp Phe Tyr Ile Ala Ile Cys Phe Val Phe Cys
             75              80                  85
```

```
TTC TGC GCT CTG TTG GAG TTT GCT GTG CTC AAC TTC CTG ATC TAC AAC        4180
Phe Cys Ala Leu Leu Glu Phe Ala Val Leu Asn Phe Leu Ile Tyr Asn
         90              95                  100
```

```
CAG ACA AAA GCC CAT GCT TCT CCT AAA CTC CGC CAT CCT CGT ATC AAT        4228
Gln Thr Lys Ala His Ala Ser Pro Lys Leu Arg His Pro Arg Ile Asn
 105                 110                 115
```

```
AGC CGT GCC CAT GCC CGT ACC CGT GCA CGT TCC CGA GCC TGT GCC CGC        4276
Ser Arg Ala His Ala Arg Thr Arg Ala Arg Ser Arg Ala Cys Ala Arg
120                 125                 130                     135
```

```
CAA CAT CAG GAA GCT TTT GTG TGC CAG ATT GTC ACC ACT GAG GGA AGT        4324
Gln His Gln Glu Ala Phe Val Cys Gln Ile Val Thr Thr Glu Gly Ser
             140                 145                 150
```

```
GAT GGA GAG GAG CGC CCG TCT TGC TCA GCC CAG CAG CCC CCT AGC CCA        4372
Asp Gly Glu Glu Arg Pro Ser Cys Ser Ala Gln Gln Pro Pro Ser Pro
             155                 160                 165
```

```
GGT AGC CCT GAG GGT CCC CGC AGC CTC TGC TCC AAG CTG GCC TGC TGT        4420
Gly Ser Pro Glu Gly Pro Arg Ser Leu Cys Ser Lys Leu Ala Cys Cys
             170                 175                 180
```

```
GAG TGG TGC AAG CGT TTT AAG AAG TAC TTC TGC ATG GTC CCC GAT TGT        4468
Glu Trp Cys Lys Arg Phe Lys Lys Tyr Phe Cys Met Val Pro Asp Cys
     185                 190                 195
```

```
GAG GGC AGT ACC TGG CAG CAG GGC CGC CTC TGC ATC CAT GTC TAC CGC        4516
Glu Gly Ser Thr Trp Gln Gln Gly Arg Leu Cys Ile His Val Tyr Arg
200                 205                 210                     215
```

```
CTG GAT AAC TAC TCG AGA GTT GTT TTC CCA GTG ACT TTC TTC TTC TTC          4564
Leu Asp Asn Tyr Ser Arg Val Val Phe Pro Val Thr Phe Phe Phe Phe
            220             225                 230

AAT GTG CTC TAC TGG CTT GTT TGC CTT AAC TTG TAGGTACCAG CTGGTACCCT         4617
Asn Val Leu Tyr Trp Leu Val Cys Leu Asn Leu
            235             240

GTGGGGCAAC CTCTCCAGTT CCCCAGGAGG TCCAAGCCCC TTGCCAAGGG AGTTGGGGGA        4677

AAGCAGCAGC AGCAGCAGGA GCGACTAGAG TTTTTCCTGC CCCATTCCCC AAACAGAAGC        4737

TTGCAGAGGG TTTGTCTTTG CTGCCCCTCT CCCCTACCTG GCCCATTCAC TGAGTCTTCT        4797

CAGCAGACCA TTTCAAATTA TTAATAAATG GGCCACCTCC CTCTTCTTCA AGGAGCATCC        4857

GTGATGCTCA GTGTTCAAAA CCACAGCCAC TTAGTGATCA GCTCCCTAAA ACCATGCCTA        4917

AGTACAGGCG GATTAGCTAT CTTCCAACAA TGCTGACCAC CAGACAATTA CTGCATTTTT        4977

CCAGAAGCCC ACTATTGCCT TTGTAGTGCT TTCGGCCCAG TTCTGGCCTC AGCCTCAAAG        5037

TGCACCGACT AGTTGCTTGC CTATACCTGG CACCTCATTA AGATGCTGGG CAGCAGTATA        5097

ACAGGAGGAA GAGATCCCTC TCCTTTGGTC AGATTATTAT GTTCTCAGTT CTCTCTCCCT        5157

GCTACCCCTT TCTCTGCAGA TAGATAGACA CTGGCATTAT CCCTTTAGGA AGAGGGGGGG        5217

GCAGCAAGAG AGCCTATTTG GGACAGCATT CCTCTCTCTC TGCTGCTGTG ACATCTCCCT        5277

CTCCTTGCTG GCTCCATCTT TCGTCTGCAC TACCAATTCA ATGCCCTTCA TCCAATGGGT        5337

ATCTATTTTT GTGTGTGATT ATAGTAACTA CTCCCTGCTT TATATGCCAC CCTCTTCCTT        5397

CTCTTTGACC CCTGTGACTC TTTCTGTAAC TTTCCCAGTG ACTTCCCCTA GCCCTGACCC        5457

AGGCACTAGG CCTTGGTGAC TTCCTGGGGC CAAGAAACTA AGGAAACTCG GCTTTGCAAC        5517

AGGCATTACT CGCCATTGAT TGGTGCCCAC CCAGGGCACA CTGTCGGAGT TCTATCACTT        5577

GCTTGACCCC TGGACCCATA AACCAGTCCA CTGTTATACC CGGGGCACTC TAACCATCAC        5637

AATCAATCAA TCAAATTCCC TTAAATTTGT ATGGCACTGG AACTTTGGCA AAGCACTTTT        5697

GACAAGTTGT GTCTGATTGG AGCTTCATGA TAGCCTTGTG ACATCTTTAG GGCAGGATTC        5757

TTATCCCCAT TTTGCAGATG AAAACCCTGA GTCACAGATT TCTGTGGGAC TGTGGATCTC        5817

ACTGGAAGCT ATCCAAGAGC CCACTGTCAC CTTCTAGACC ACATGATAGG GCTAGACAGC        5877

TCAGTTCACC ATGATTCTCT TCTGTCACCT CTGCTGGCAC ACCAGTGGCA AGGCCCAGAA        5937

TGGCGACCTC TCTTTAGCTC AATTTCTGGG CCTGAGGTGC TCAGACTGCC CCCAAGATCA        5997

AATCTCTCCT GGCTGTAGTA ACCCAGTGGA ATGAATTTGG ACATGCCCCA ATGCTTCTAT        6057

ATGCTAAGTG AAATCTGTGT CTGTAATTTG TTGGGGGGTG GATAGGGTGG GGTCTCCATC        6117

TACTTTTTGT CACCATCATC TGAAATGGG                                          6146
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 242 amino acids
    (B) TYPE: amino acid
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Val Met Thr Ile Phe Phe Asn Val Ser Arg Arg Phe Gly Tyr Val
 1               5                  10                  15

Ala Phe Gln Asn Tyr Val Pro Ser Ser Val Thr Thr Met Leu Ser Trp
                20                  25                  30

Val Ser Phe Trp Ile Lys Thr Glu Ser Ala Pro Ala Arg Thr Ser Leu
            35                  40                  45

Gly Ile Thr Ser Val Leu Thr Met Thr Thr Leu Gly Thr Phe Ser Arg
        50                  55                  60

Lys Asn Phe Pro Arg Val Ser Tyr Ile Thr Ala Leu Asp Phe Tyr Ile
65                  70                  75                  80

Ala Ile Cys Phe Val Phe Cys Phe Cys Ala Leu Leu Glu Phe Ala Val
                85                  90                  95

Leu Asn Phe Leu Ile Tyr Asn Gln Thr Lys Ala His Ala Ser Pro Lys
            100                 105                 110

Leu Arg His Pro Arg Ile Asn Ser Arg Ala His Ala Arg Thr Arg Ala
        115                 120                 125

Arg Ser Arg Ala Cys Ala Arg Gln His Gln Glu Ala Phe Val Cys Gln
    130                 135                 140

Ile Val Thr Thr Glu Gly Ser Asp Gly Glu Glu Arg Pro Ser Cys Ser
145                 150                 155                 160

Ala Gln Gln Pro Pro Ser Pro Gly Ser Pro Glu Gly Pro Arg Ser Leu
                165                 170                 175

Cys Ser Lys Leu Ala Cys Cys Glu Trp Cys Lys Arg Phe Lys Lys Tyr
            180                 185                 190

Phe Cys Met Val Pro Asp Cys Glu Gly Ser Thr Trp Gln Gln Gly Arg
            195                 200                 205

Leu Cys Ile His Val Tyr Arg Leu Asp Asn Tyr Ser Arg Val Val Phe
    210                 215                 220

Pro Val Thr Phe Phe Phe Phe Asn Val Leu Tyr Trp Leu Val Cys Leu
225                 230                 235                 240

Asn Leu
```

(2) INFORMATION FOR SEQ ID NO:3:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 556 base pairs
        (B) TYPE: nucleic acid

```
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)



    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TCCCCGGCTC CCCACCGCAG CAGCCGTCAC GTCGTCGGAG ATTTCCATCG GGGCGGGCCT        60

GGGGCGGGGA GCGCGGGACG GGGCGGGGCG AGTGGGAGGA GTGAAAGTTG GAGCCCAGCA       120

AAAGCCTCCG CCCCGCGCTC AGTGCGGCCA GAGCGTGAGC CGCGACCTCC GCGCAGGTGG       180

TCGCGCCGGT CTCCGCGGAA ATGTTGTCCA AAGTTCTTCC AGTCCTCCTA GGCATCTTAT       240

TGATCCTCCA GTCGAGGTGA GTCTCCATCC CGGGACCCGG GAGCCCTTCG CGCCCAGCTC       300

CCTCTCCCCG GGAGCCGGGA CGGCTCCCGG GACCCCAGCG GCCCCGCGTT CCTCGAGCCC       360

CGCGCCCGCT TTGCCCCGGC CCTACCGCGG GCTGGCCGAG TCCCGCGTCC CCTCGATGCG       420

CGCCGGCCTC GGCCCGCCTC ACTGTAGGAT GGGCTCCCGG GGTCCTTGAG GGGGAGCTCC       480

AAAAGGAAGA CAGGACGCCA GAAGGAAGAC GGGACTCCAG TTCGCGGATT CCCGCTCTCA       540

AAAGCACTGC GGTGGC                                                      556
```

```
(2)  INFORMATION FOR SEQ ID NO:4:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 1097 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)



    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GGCACTAGCT GGGGCCCCTA CAGAGTGCAG GGCAGAGCTT CATTTTTCGT TTGAATGTTA        60

TAGGGTCGAG GGACCTCAGA CTGAATCAAA GAATGAAGCC TCTTCCCGTG ATGTTGTCTA       120

TGGCCCCCAG CCCCAGCCTC TGGAAAATCA GCTCCTCTCT GAGGAAACAA AGTCAACTGA       180

GACTGAGACT GGGAGCAGAG TTGGCAAACT GCCAGAAGCC TCTCGCATCC TGAACACTAT       240

CCTGAGTAAT TATGACCACA AACTGCGCCC TGGCATTGGA GGTGAGGAGC AGAACGACGT       300

TCTTCCCCTC CTAGAGGGTC CAGGGGTTGA GGGCATAGGC ATGGAGAATG CACCTGGGCA       360

GTAACAGAGG GTGCCATGCT CATGGACAGG AACATCTGCT ATTGACCTGT CAGGTAAGAG       420

ATATTAACTC TATTCTCAGC AGTGTCATTG ACCTTGATCA AGACTTTTCC CTTCTCTCGC       480

CCTCAGTTTT TCCAGTGGTA AAATGAGAGG ACTAAACTAG ATTGTTGATC TTCAAGATGT       540
```

```
GTGTCCAATT CTTAACAGTC CGTGAGCTTG GTTTTGCCAT GAAAGAATAA ATAAAGAAAT    600

AGGATTAGAT GCTGAAACTG TGTGGTCCAA CACTTACTTG ACTCCCCTTT CATCCCCTCT    660

GACCACTTCC TCCCCCGTCC CATGCGCCTG TTTGACACTT ACCCTCTGCT GCTTCTGCTT    720

CCCTTATAGA GAAGCCCACT GTGGTCACTG TTGAGATCTC CGTCAACAGC CTTGGTCCTC    780

TCTCTATCCT AGACATGGTG AGTACTAAGC TTTTTTAGTA CTATTTCCTA GCCCAGGGGC    840

TGACCTATGG GACCTTCCAC AGACTTCTGC TTTCTGCTCT GTACTTCTGT AACAACTCCA    900

ATAATTTATT TTCTGGAGGG AGAAAGGGAT TTTTTAACCA CTGGTTTGAG AATGAGACTG    960

GAAAAGGTAA GTCCCTTGCT ACTTGAAGAG GATCTTCAGA ATCATGACCA TATCTTCCAG   1020

TTTTTTCATT CAAAATAGAA ATAATAAAGC AATGTTAAAC ACCACAATGG ACTGCCTCTT   1080

TTCTTCTGAT AATTATT                                                 1097
```

(2) INFORMATION FOR SEQ ID NO:5:

    (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 9631 base pairs
       (B) TYPE: nucleic acid
       (C) STRANDEDNESS: double
       (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
TGCATGATGT GAGGACCTAG CTCCCCTTGC TCACGGTAAT CAAGCAGGGC ATGACTGACT     60

TCTCATTTGG GATCATTAGG CTTGGTAACC TCTGAGTTCC CTTCTAGTTC TGACATTGGA    120

AGAGTCTTTG AGCAGAGAGA AGGGACCTCA TCTGACTCCT GCTTCCTGGT CTAGTGCTCA    180

ATGCATTCCA CCACATTGCC TTTAGGATTC TAGGTTAGGG AGTGGCAAAC AATATGTTTG    240

CATGTGAATG TCCTTTTTCT GTTCATCCCC AAGTATGTGC TTTTCTGTCC TTCCCACCAG    300

GAATACACCA TTGACATCAT CTTCTCCCAG ACCTGGTACG ACGAACGCCT CTGTTACAAC    360

GACACCTTTG AGTCTCTTGT TCTGAATGGC AATGTGGTGA GCCAGCTATG GATCCCGGAC    420

ACCTTTTTTA GGAATTCTAA GAGGACCCAC GAGCATGAGA TCACCATGCC CAACCAGATG    480

GTCCGCATCT ACAAGGATGG CAAGGTGTTG TACACAATTA GGTATGTCAA GCCTCTGGAG    540

TCTCACTTCC TGGAATTCTC TCTCCCCTTC TGATAATTTT AGCTAAAGAT CCATGGGCAG    600

AGATCTCATC CTGAATGATA CCTCTAAGGG CCTGTCCAGC TTTCCTAGAC CATGAGCTCA    660

GCCCCCTTAT GTAACAGATA TAGAGGCCTC AAAATAGAAA GATATTGCTT AAAGCCACAC    720

ACCAAGTTTG TGGCAGAGCT GGAACTGGTA CTCAGTTACT TGGCTCCGAG TCCAGAGCTC    780

CCTCAACTAG GATGTGCCAG TATGACTGCA TTATCTAGAC AATTCCATCC TACGTGGGCA    840
```

```
CTCGATACAA AGATACGTCC ACAGTGGTGG AATTGTTCAG GCAGAGCAGC AGCACGTAGT      900

GGCAAAGGTA CCTAAGATCA AGTTGGATAC TTGAATTCCC AGCAGGGGAA GGTTGTGTGT      960

GGGGATAGCA GGGAGGATGT TGGCAGTTCC TGGAAACTAG GGTGGGCGAG AAAACAAAAG     1020

CCGATCGAAG TTGCTCCATA CGTTTCTCTA ATGATGGAGC CCAGAGTAAC CAGATACTTC     1080

TAAGCTGTTT GTTTGTTTTG TTTTGTTTTG TTTTGTTTTG TTTTGTTTTG TTTTCTCTCT     1140

TGTTATCTCT CCTTTGAGCT TTTTGTCTTA AATTCTAGCG AGGTCCAGGC ACGGTGGCTC     1200

ACGCCTGTGA TCCCAGCACT TTGTGAGGCT GAGGCAGGCA GATCACTTGA GGTCAGGAGT     1260

TCGAGACCAG CCTGGCCATC ATGGGAAAAC CCTGTCTCCA CTAAAAATGC AAAAATTAGC     1320

AGGGTGTGCT GGCACTAATT CCAGCTACTC GGGAGGCTAG GGCATGAGAA TTGCTTGAGC     1380

CTGGGAGGCA AGAGGCTGCA GTGAGCTGAC ATCACGCCAC TGCCCTCCAG CCTGGGTGAC     1440

AGAGTGAGAC TCTGTCTCAA ACAAACAAAG AAAAAAATTG ACTCTGGCCA TTCATTGGTG     1500

GTAGTCCCTA GACCAAAGCT GGGTGGATAC GGAAGTGCTT AGGGCCAGCC TGATGAGGCT     1560

CCTTTCTCCC TTCCAGGATG ACCATTGATG CCGGATGCTC ACTCCACATG CTCAGATTTC     1620

CAATGGATTC TCACTCTTGC CCTCTATCTT TCTCTAGCTG TGAGTACCTT CTTAAGTTTC     1680

TGGGGCCCCA GAAACATGCT GGGCTCCTTC TTTTTCTCAT CCTTGCCATT TACATTTTTC     1740

TGCCTCTGCT TTTCTTCTAA AATGCTGCCA AGGTTGTGCA GGACTTCCAT CCTCCACCCT     1800

CATTTCCTTT CCTGCCAACA ATACTGTGTT GCTCATCCCT TCCACGTGCC TCTGAAGCGT     1860

ATCTCAAGTA TGTCTGCTCC TCTCCATCTC CACTGGCACT ACCTTGGTTT AGGCCTTTGT     1920

TATCTTCCAC CTGGACTTTT GCCACATCTT CACTTTGAAA CTGCACATGT CCAAAATGAA     1980

ATTCATTGTC TCCTCCAAAC CTCTACCACC AAAACAAGTG TGTTGCTTCT GGGTTCCCAT     2040

CTGTCTCATT GAAGAGGACC ATCACTCACC CAGTTGCGCA AATCAAGAAC TTTGATGTTT     2100

CCTCTCCCTC ACCTCCTGCA TCTAATCAAT CAGCACATCC TGTTGGTGTT CCTCCCAGT     2160

CTCTATCGAT GCTGTCTATT TCTCTGCACC CTGTACAGCT TTGACTTCCA CCTGCATTAA     2220

TTTAATTCTG CCTGGATTAC TACACTGGCC TCCTTGACAA CATGTTGTCC TCACAGAAGG     2280

ACCAGAGTGA CCTAGCTGAA GGGTCACCTA GGTTGGGTCA CTTCTTAGTC TCGAATCTGC     2340

CGTTAACTCT CATGGATCAA TTTGAAATTC CTTAGAATGA ACCTCAAGGC CATTCATGAA     2400

CTGGACCCTG CCACCCAATC CTGTGCACCT CATCCTCTGT GAGCTAGCCA TCCTGAACTT     2460

TTGTCCTTTC CACAATACAC CAGGTGTTTC ACCTTTCTAT ACTGCCCCTT AACCCCTTCA     2520

ACCTCATTCT TATTGAGAAT ATTTACTTGA GTTTCAAGAT TTAATGGGAA TATCACCTGC     2580

TTTATGAAGT CTTTTCTGAG TATGTCCCCA AGTGACCTTT ATCTACTTTG TTTCCCCGCT     2640

GTTCTGTGGA CTTAGGTTTT TCAGAGCTCC TCCAAAAATC ACAGTAGTAT ACTCACTGTC     2700
```

```
TTATAAAATT AAATGTGATT GCTTGAGGGT AGGGTTCATG CCTTGCTCAT CTCTGTATTT    2760

CTGGCCTAGG GCCTGATACT GAGGAATGCT CAGTAAACGC ACTCATTGAA TGGACTTCAA    2820

CAATGAGGTA AGAGAGGCAA GGTCCCACAG CTGGTGAGGC CAGAGACAGG ACTCCAAGGC    2880

ATTGTGCAGG CTGAGTTCAT GCTATTGGAG ACCTCAGGTG GGCTTCCAAG TCTCATAGGA    2940

CCCTCTTTCT CACATTCCTT TCCAGTTTCC TATCCTGAGA ATGAGATGAT CTACAAGTGG    3000

GAAAATTTCA AGCTTGAAAT CAATGAGAAG AACTCCTGGA AGCTCTTCCA GTTTGATTTT    3060

ACAGGAGTGA GCAACAAAAC TGAAATAATC ACAACCCCAG TTGGTAAGCG TGCCAGGGCT    3120

TGGCGGAAGT CCAGGAAGGT GGTAGGGATG ATTGGAGATG GCCATCCATA CAAATGCTTT    3180

GCAGTCATCC CGTGCAAACA TTGTAAGACA TGGCTCCTGT CTTATAATTG CTAAGCACTT    3240

ACAACTGTTT GCAGAGGAAA CTGAGACTTT GTAACTATGT CTCAGTCTCA TCTGCAAAGA    3300

AGTAAGTGCT TTGCCAAGCT CCTTGAGAGG TTAGGTAAGT AGATAAAGTT CTGCTGCTGT    3360

CGGAATGTGC AGCTGGCTTT TTCATGCAGA CCCTTCAGTT TCGAGGTTAC AACTCTGACC    3420

TCTTTGGATG ACTTTGGGGA ATGGAGCTCG TGTGAGTTCT CCATACCCAG AACCAATCCA    3480

GTCTGGTTGA ATGGGAAGCA AAGTCCATTG TAGTGGGAGG TGGAGGCTAG AGTTCTAATG    3540

TCAGCTAGTT TAAGGCTGGG AAAGTCTGGA GGAAGTTACA GCAGCTACAC TGGCTGCTGC    3600

ATTGACATTT ATCTTAAAGG AACAAGTCTG AAAAGCACAG ATTCTTATCA AAGGCTTCAT    3660

GGTGGATTCC ACATAGACAT AGTGGCCACT GGTTTTCTGA CCTTTTCTCT GACAAAGACT    3720

AAAGGGGAAG GTCCTGGGTA TCTTACACTT CAGCTCCCAA TTAGATGTGA GCACCTTCAC    3780

TTATGTTCCT AGGTGACCTG AATGAGGAGC CAAGGGACCT CCCCAGGGTA GCTCCCAGAG    3840

CAACCCTGGA AACACTCTTC ACACATCCTG ACCAAGTTCA GGGCAGTGAA GGCACTGCCC    3900

TCATCGTTTC CAGAATGTGG ATGGAGCCAG TCACCCAACC AGCCATTTGT CGTGAGAGGC    3960

ATCTTGTTCT GCTACCATGT GACTAGGCAG AAAATCTGCT TTTGTTTCAT TTATTGAGTC    4020

AGTCTCTGGA TGAGGGAAAG CTCATGCTCA TGTGGCTAGA GCTTTGCTTG CACAGTATTA    4080

GGCAGGGGCA GAGGGCTGGG CTACCTTAAA AATACTTGCC CTTTTTCTTG GGGACTCTGG    4140

GGAAGCGGTT TTACTACCTT TGACTTGGGA GCCTTGCTCT TCTGCCAGCT AACCATGGGC    4200

CTGCCTCTTG GTTTTCTGCA CCTCAGCTTT TCCCGGATAG GTGGGGACCC ATCATCAAAA    4260

GTGACAGAGA AGATAAGGCC CAGGGGCTTT CAAGTCACTA GTGGTTCCGT TTAGTAGATG    4320

ATTGTGCATT GTTTCAAAAT GGTGCCCTAG TGACTACAAA GCCCCAGAGC CAGCATCATC    4380

ATCAAAGCAA TGACAGTAGG TAAGCACCAG ACCTCCTTGG GAGTGAGGAG GATTCTTGAG    4440

GAGAAAAGAG GTCTTCTTTC TCCTCTGCTG GAGACTAGTT GATCTGGAGA CGTGGTTCCT    4500

TCAATGTCAG AGTTATCTTT GGGACTGGTC TCAAACTCTT CCAGTTGGGC CCTGGGGCAG    4560
```

```
GTCTCTCCAT CTGGAGCATA CTTACGTGCT CGGCGATTAA GGGTTCAGAA TGCAGTGGTA    4620

GCCTGCTACT CTGGCCATCT TGGACCTTGA TCCAGAGAAT CTCTGCTTCA GGAGCTTCTA    4680

AGAGAGTCCA GCCCTGCCTC CAGAGAGAGG CTTGCCCTTC ACTGATGGCT GTGGAGCCTC    4740

TGATGGAATA TTATTGCTGG TCAGGAATTC ACTGTCTTAC AAGGAGGTTT CCTTCTTCTC    4800

TAGACAGTTC TGTTCATCAA AAAACTCTCC CTGTTCTTCT GAAATTGGAG TCTCTGGAAG    4860

TTCCACACAT TAAGCTTAGT TCTTTTTCCT TGGAACTGTC CAGGTTACAT TAGTCCAGCC    4920

ACTGTTTCAC AGGACCGAGA TTAAACGATC AACATCATCA TTCCCGGCAT GGATCATAGT    4980

CTGTTGTAGT CTACATAGCC CTAGTTTATT TTTCTTCCCT TATTCTTCAA AGCTTTGGGT    5040

CCATTCATTC TTCTAGTCCC AGTCCTCTGG ACATGGTCTA TTTAATTGTG TCCCTCTGAC    5100

ACTGCAGTGA CCAACCATGA TCTGGTCAAA GAGGATAAGA GTTTGAGCAG AAAACCATCT    5160

TTAGCATATA TTTTTTTGCT TTGGTTCATC AGCCCCAGAT ATATTGTTTT CCTTACCCGT    5220

GCTTCTCTCA CTCCTCAAGA AGAAGAAAGT GTGTGTTAGC ATCTTTCTCT TGTCCTTCAA    5280

GACAAATTGG CATCTCTTGA CGAGCGGAGA AGGTTCTTTT TTGGCCAGAA TAAATAAAAT    5340

TAAAATAGAA TCATCCAACA GAATAATAAA TCTTCGTGCA ACAAGAATAT ATTATATAAA    5400

CCCAGCAATT TTGCAGGGCC TGGGTATAAC TAATTAGAAG TGTCTTAAAT TGCAGTCAAG    5460

ATCCCACGGC AAGAGGACTT TTGATAAATA CATTCTGGCC AGTAGGCAAG TGCGAGGGTG    5520

GTCCGTGCAG CAGCTCTGGA GGAGTTCTAT CCCAAAGCTA TACTCAACAC ACAGGTTTCC    5580

CACTGACAAC AGGTCGCTCC CTTGCCTTCT TCCAGAAGAA TCTGAGAAGC TTTGCTCCTT    5640

GAGTTTCAGT GCTGCCAAGG TGAGTACGAA AGGCTGCTCT TCTCATTCAG CTCCAGCCCA    5700

CCCAGACCTG CTGGGCAGTT GATCCACTTT CCAAAATAGG AGGACACACG GACAGGTTAG    5760

TGTTCTGGTC TGCTTTACAA AGCTGTTGCC TGACAGGAGC AAGAGTTGCT GAGTGTCTGC    5820

TGGGTTCCAG GCTGTTCTGA GCTTGGATGG GCAGGGGCTA AGCCACAGGG CCTGCATGAG    5880

CCCTGCCTTG AAGGGACTTA AAAGACGACC TAATTATAGG CCTAGGAATT TTACAGTATT    5940

GCAACTGCAA TGTGATGCTG AAAGTGGAAA ATGATGTCCT GGGCTCAGAG AAAAGCCCAC    6000

ACCAGCCTGG GAGTCATGAT AGCAGCAGAG TGCTTGGGGA GGGTGTGTCA GAGCATAAAG    6060

CAGCATGAAT GCTACAAAAG AAGATGCCAA CTAGAGATAT AGGTTGTCAT CAGGTCCCGG    6120

AGGAGCCATG ACCGTCTAGC TGAGAGCCAT GACCAAGGAC ACAATGTCCA AGTGACTGTG    6180

AGGACCTCAG TCTGCCCTGT GGATGTGTAT GCCACAGACC TGACTTCTGG AGGGCTGACT    6240

GAAATGTTCA TTTTAAGCTT TTTCTTCTCT TTCCCTGAAA CACTCAGTTT GGGTTAGGGG    6300

TCATAGACTA AGACCAAAGA GTCCAGGGTT AGAATCTTGG TGTAAAATTG CAGGCCATCT    6360

CAGGAAATCT GTGAGCAGAT GGGATTGGCT TTGGGTAAGG TGCGTGTGGA AAATGTCAGT    6420
```

```
GGGAGCCGGG TCATGGTGGG CCTTTAGCAT CAGATTCCAG AGTGCAGATA GTCTGTATAG    6480

CTCATGTGAA ACAGGGAGCC ACCAAAACTT TGGGGAGCAG GCTAGTGCCG GTTTTGACCA    6540

CCTGTGGAGC AGTGCTCACT CACGAAGGCA TTTTGCCATC ACATGAATGT GCAGAAAGGA    6600

GGCCAAAAGC ATTCTGTGCT TCTCCACCAC AGCACAGACT TCCCTAGTCT CATTTGCTGA    6660

GAGTAGACAT TCTGAGGGCC AGCAGTGCAG GTGTGATGTG CCTCAGAGGG TATGAAGCCC    6720

TTAGTCAGCC ATCTGGATAT CAGCTGCGTG GGCATGATAT CTAGAAGGCT AATTGATTTT    6780

TTCACTTTCA CCTGACTCTC TTGCCAACCT GCAGAGACAG ACATTGGGTG TAGGACAGTG    6840

AACTGAGAAG GAAGCTATTA AGATTCTGGC CTTGGCTTAG CTCTCAACTG GCCATTGGTC    6900

TTGCAGTAAG TCTTTTTTCT GGGCTTCTTC TGGTCCTATT TGTATGTATT GCATTGTCAC    6960

ATCATGCCTC TATCCTAGGG AATACTGTGA GCTGAAAAAT GAGACCCTTA CTGTTCACGT    7020

CCTGCTAAGG GGGACCGTCG TGTCAGCACT GTAATGCAGT GATGTTTTTT GTGTCTTTCA    7080

GGTGACTTCA TGGTCATGAC GATTTTCTTC AATGTGAGCA GGCGGTTTGG CTATGTTGCC    7140

TTTCAAAACT ATGTCCCTTC TTCCGTGACC ACGATGCTCT CCTGGGTTTC CTTTTGGATC    7200

AAGACAGAGT CTGCTCCAGC CCGGACCTCT CTAGGTAAGA GGAGAAACAG GTATACGCAT    7260

AGGCACATGG CTGGGAGTTG GCTGGGCCAG GGCAGAGTTG CCTTGTCATG GAGTCTTTTA    7320

ACCAATGTCG CACATAGGTC AGGAGCTGAG CCCATCACTC TTGTGCTCTT GCAGGGATCA    7380

CCTCTGTTCT GACCATGACC ACGTTGGGCA CCTTTTCTCG TAAGAATTTC CGCGTGTCT    7440

CCTATATCAC AGCCTTGGAT TTCTATATCG CCATCTGCTT CGTCTTCTGC TTCTGCGCTC    7500

TGTTGGAGTT TGCTGTGCTC AACTTCCTGA TCTACAACCA GACAAAGCC CATGCTTCTC    7560

CTAAACTCCG CCATGTATGA GCTGGGTATG GGAGTGGTGG CAAGGCTTTG GAGTGTAGAG    7620

ACATGCTAGC AAGGGTACTG GGGTTATGGC ACATGGGTGG TCAGCTTGCT GAGTGATGGA    7680

ATGTTACCCA GGGTGGTGGC GGGGTTGAAT CAACTTCCTG ATGTAATGGT GAGAAGTTGG    7740

AGGAGAGAAG CCAAGATATG GTGTGCCAAA GACAGTTTCC AGAAAATCCG GAGGCAGCAC    7800

TTAGACTTGG GTTATCTTCC CTTGACTTTT CCCCACTTCT TTCCTTGTCC ATTTTAGCCT    7860

CGTATCAATA GCCGTGCCCA TGCCCGTACC CGTGCACGTT CCCGAGCCTG TGCCCGCCAA    7920

CATCAGGAAG CTTTTGTGTG CCAGATTGTC ACCACTGAGG GAAGTGATGG AGAGGAGCGC    7980

CCGTCTTGCT CAGCCCAGCA GCCCCCTAGC CCAGGTAGCC CTGAGGGTCC CCGCAGCCTC    8040

TGCTCCAAGC TGGCCTGCTG TGAGTGGTGC AAGCGTTTTA AGAAGTACTT CTGCATGGTC    8100

CCCGATTGTG AGGGCAGTAC CTGGCAGCAG GGCCGCCTCT GCATCCATGT CTACCGCCTG    8160

GATAACTACT CGAGAGTTGT TTTCCCAGTG ACTTTCTTCT TCTTCAATGT GCTCTACTGG    8220

CTTGTTTGCC TTAACTTGTA GGTACCAGCT GGTACCCTGT GGGGCAACCT CTCCAGTTCC    8280
```

```
CCAGGAGGTC CAAGCCCCTT GCCAAGGGAG TTGGGGGAAA GCAGCAGCAG CAGCAGGAGC      8340

GACTAGAGTT TTTCCTGCCC CATTCCCCAA ACAGAAGCTT GCAGAGGGTT TGTCTTTGCT      8400

GCCCCTCTCC CCTACCTGGC CCATTCACTG AGTCTTCTCA GCAGACCATT TCAAATTATT      8460

AATAAATGGG CCACCTCCCT CTTCTTCAAG GAGCATCCGT GATGCTCAGT GTTCAAAACC      8520

ACAGCCACTT AGTGATCAGC TCCCTAAAAC CATGCCTAAG TACAGGCGGA TTAGCTATCT      8580

TCCAACAATG CTGACCACCA GACAATTACT GCATTTTTCC AGAAGCCCAC TATTGCCTTT      8640

GCAGTGCTTT CGGCCCAGTT CTGGCCTCAG CCTCAAAGTG CACCGACTAG TTGCTTGCCT      8700

ATACCTGGCA CCTCATTAAG ATGCTGGGCA GCAGTATAAC AGGAGGAAGA GATCCCTCTC      8760

CTTTGGTCAG ATTATTATGT TCTCAGTTCT CTCTCCCTGC TACCCCTTTC TCTGCAGTTA      8820

GATAGACACT GGCATTATCC CTTTAGGAAG AGGGGGGGGC AGCAAGAGAG CCTATTTGGG      8880

ACAGCATTCC TCTCTCTCTG CTGCTGTGAC ATCTCCCTCT CCTTGCTGGC TCCATCTTTC      8940

GTCTGCACTA CCAATTCAAT GCCCTTCATC CAATGGGTAT CTATTTTTGT GTGTGATTAT      9000

AGTAACTACT CCCTGCTTTA TATGCCACCC TCTTCCTTCT CTTTGACCCC TGTGACTCTT      9060

TCTGTAACTT TCCCAGTGAC TTCCCCTAGC CCTGACCCAG GCACTAGGCC TTGGTGACTT      9120

CCTGGGGCCA AGAAACTAAG GAAACTCGGC TTTGCAACAG GCATTGCTCG CCATTGATTG      9180

GTGCCCACCC AGGGCACACT GTCGGAGTTC TATCACTTGC TTGACCCCTG GACCCATAAA      9240

CCAGTCCACT GTTATACCCG GGGCACTCTA ACCATCACAA TCAATCAATC AAATTCCCTT      9300

AAATTTGTAT GGCACTGGAA CTTTGGCAAA GCACTTTTGA CAAGTTGTGT CTGATTGGAG      9360

CTTCATGATA GCCTTGTGAC ATCTTTAGGG CAGGATTCTT ATCCCCATTT TGCAGATGAA      9420

AACCCTGAGT CACAGATTTC TGTGGGACTG TGGATCTCAC TGGAAGCTAT CCAAGAGCCC      9480

ACTGTCACCT TCTAGACCAC ATGATAGGGC TAGACAGCTC AGTTCACCAT GATTCTCTTC      9540

TGTCACCTCT GCTGGCACAC CAGTGGCAAG GCCCAGAATG GCGACCTCTC TTTAGCTCAA      9600

TTTCTGGGCC TGAGGTGCTC AGACTGCCCC C                                     9631
```

(2) INFORMATION FOR SEQ ID NO:6:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 7479 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
TTCTACCAGA CCTGGTATGA TGAGCGCCTC TGTTACAATG ACACCTTTGA GACACTTATT        60
```

```
CTACATGGCA ACGTGGTGAG CCAGTTGTGG ATCCCGGATA CTTTTTTTAG GAATTCTAAG    120

AGGACCCAAG AGTATGATAT CACCATACCC AACCAGATGG CTCTCATCCA TAAGGATGGA    180

AAAGTGTTGT ACACAGTTAG GTATGTCAAG CCTTTAGTGT CTCACTTTCT AGGGCTCTCT    240

CACTCTTCCA GAAAATTTTA GCTAGGGACT CCTAAGTAAA TATATCATTC TGAATAATAT    300

TCCTAAAACC CTGTGCAGCT TTTCTAGATC AGGAATTCAA CCTCTTCCTG TGCAAATATA    360

AGACACTGAG ATATTAAATA AAAGATTTTA AAGATGCATG CCAAGTGTTT TGGTAATCAT    420

GGAGCTGTAT TTGCCCCACA GTCTAGAAGG TAGGTGCTGG CCTGACTGTA TTAGTAGATA    480

ATTCCATCCC AAATGGGCAT AGTATTTAAA GACACGCCCA CAGTGACAGA ATTGTCCAGG    540

CACAGTATAA ATACATCGGT GGAAAAGTAT TCAGGACTAG GATACATAAA TTGAATGCCT    600

GGAAGGAGAT GCTTGTGTCA GCAATATGGG ATGTGGAGGT GGTGCAAAGT AGCACAGGAC    660

TAGAGCACTG CACGAACAAG AAAGAAAAAT CAATCTAGGC TGGCCCATGT ACTTCTTTGA    720

TGATGTAGCC TAGTAGTCAG ATTTTAGAAG TTCTTCTTCC TCTCTTTGTA TCATTCCTTT    780

GGGCTTTTTG TCTTAAATTC TACTAATGGC ACTCATTAGA TATGATACCT GGTCAATGTT    840

GAGTAGATAT CAGTGCAAAG AGCCAGCCTA ATGAGGCTCC TTTCTCCCTT CCAGGATGAC    900

CATTGATGCA AGATGCTCAC TCCACATGCT CAATTTTCCA ATGGATTCTC ACTCTTGCCC    960

TCTGTCTTTC TCTAGCTGTG AGTACCTTCT TAGGTTTCTG GGGATCCAGA GATATGCTGG   1020

GCCCCACTTT TGCTCATCTT TGCCTTTTAC ATTTTTCTGC TTCTGCTTGT TTCCTGTGAT   1080

GGTGCCAGGG TTGCTATAAG CCTCCATTTT CCATCATCAC CCCTTTTCCT GTAAGCAGTG   1140

CTGTGTGTGT ATTGCTTTGG CTTTATGTTG AATTCTTCCC TCACCCTTGC TCTACCTTAA   1200

GCTAGTCACC TGCTTCTGAA GTATATCTGA AGTATGCCTG CTCCTCTCCA TCTCTGCCTT   1260

GATGCAGGGC TTTAATCATC TTCTGCTTGG ACTTCTACAG TACCTTCACT TTGAAACAGA   1320

GTGTGTCTTA AGCTAAATTC ATTTTTTTCT TCCCAACACC TGCCTCCAAA GCAAATGTGT   1380

CGCTTCTGTG TTGCCTCATT GAAGGGTGCC ATCATTTACC CAGCTACCCA AGCCAAGAAC   1440

TCTGACAGCC CCATTCATTG ACCTCCTACA TCTACTCAGT TCACACATCC TGTTGGTGTT   1500

CCCTCCAAAT CTCTATTGAT CTCACCTCCA TCTCTGCATT CTGACTATTA TTAGCTTTCA   1560

TTCCTCATTT CAAATGTGTT CTTTGCTCGC CTGGATCAAT TCTAAATGCC TTAAAATGAA   1620

CCTCAAATCC CTTCATGAGC TGACCCTGCT GCTTCCTGCT ATACAACCTA TCTGTGAAAT   1680

AGTCTCCCTA AAGTTTTGCT GTTTCCACAT TCACCTTTCT GTGCTTCCCT TTCACCTTCT   1740

TCAACCTCAT CTCTATTGAG AATGTTTATT TTCAAAACTT AAAGTGAATG CCCCATGCTT   1800

TAGGAAGTGG GTCCCAAGAG ATCTTTCCCT ATTTTGCCAA TTGTTCCATG GGTAGAGATT   1860

TTTCAGAGCA ACCCCCAAAC TCACAATTGT ATATGGCCTT ACTCAATGCA ATGTGAATGC   1920
```

```
TGGAGGGTAG GTTCACGCCT TACTCATCTT TGCATCCGAA GCCCAGGGCA AATGTACTTA    1980

TTGAATGCAC TTCAACAACT GGGCCAAAGG CAGGGTCTCA AAGCTGGTGA AGCCAAAGTC    2040

AAAAGTCCAA AACATGGCTC AAGCTAGCAC CATGCTTGTG GGGCCCTCAA CTGGGTTTCC    2100

CAGTCTCATT CATTCCTCTT TTCCACATTG TTTTCAGTTT CCTATGATGA GCATGAGATG    2160

ATATACAAGT GGGAGAATTT CAAACTCAAA ATCGATGCGA AGAACACTTG GAAGCTATTG    2220

GAGTTTGATT TTACAGGAGT GAACAACAAA ACTGAAATCA TCTCCACCCC AGTTGGTAAG    2280

TGTGCAGTGG GAGCTAGAGC TGGGCAGAGT CTGAGGGTGT AGTACCAACA ACTGGAAGTG    2340

GCTATTTGTA CCAATGCTAT GTAAAGAGCT TATGGAAACA CTGTCTAGGA CATCGCTTCT    2400

CTCTCATATA TGCTAAACAC TTACAACTGT TTGCAGAGGA AACTGAGACT TCATAGCTAT    2460

GTCTCAGTCT CATCTGCAAA GAAGTAAGTG CTTTGCCAAG CCCCTTGAAG GACGGGGTAA    2520

GTAGATATAT GTTTGTAACC TTCAGTATGT TAATCTACCT TTTGAGTTTT AGATCTTTTG    2580

TATTTAATTT CTTTCTTATA TCACTGACTA ACTTAGACAA CATTGGGGAG GGGTACTCTG    2640

TGTGATTTTA CCAAATCCTG AACCAGGTCA AGTTGGTTGA GTGGGAAGGA AGACCACCCT    2700

AGCAGGAGGT AGAAGTGGAA ACTAGAGTGA CAATGGCGGC TAGTAGGGAT TGAGGAGTCT    2760

GCAGATGGTT ACTCAGGTGC AATGGCTACA GTAATATTAT CTTAAAGGAG CAACTGTGAA    2820

GAGTACATAT TTTCACCAGA GGCTTCAGAG TGGAATCAAC ATAAACATGT TTTTTCCTCC    2880

TGATCTTGTT CCTGACAAGG ATAATAGGGG CAGATTCTGG GTCTCTTAAC TTGTATCTCC    2940

CAAATAGATG TGAGCATCTT TGCTTGTATT CCTAGGTGAT CTGAATGAGG AGCCAATAAA    3000

TCTTCTCAGG GTACCTCTGA GAGCAACCCT GGACCTTCTT CACATGTACT GACAAGAATG    3060

TGGATGGATC AGGTCACCCA GCCAGACTTT TGACATTTAT TTTGCTTTGC TTCTGTGTGA    3120

TGATGACATA AAAAAATCTG CTATTGCTAC TGTTGGTCAC TCGGTGAGGG GCAGCTTATG    3180

CTAACCCAAC TAAAGCTTTG CTTATACAAT ACTAGGCAGG GATGGGTGCT GCTTGTCCTG    3240

TTCCTGGGAC TCTGAGAAAG CCTTCTTGCT TGCTTTGACT TGGGGTTCTT GTTCTTCTTC    3300

TACATAAATG TGGGCCTGCC TCTTAGTATT CAGTTTGTCC AGATAGGTGG AGCCTCATCA    3360

TCAGAAGTGA CAGAAGAGAC AAGGTTCAGG GGCTTTCAAG TCACTAGTGG TTCCGTTTAG    3420

TAGATGGTTT TTGCATTGTT TCAAAATGGT GCCCTAGTGA CTACAAAGCC CCAGAGCCAG    3480

CATCATCAGT GAAGCAATGT CAGTAGGTAA GCAACATCTC TTACTGTGGA GAAAGAAATA    3540

GGTCTTCTTT TTTCATTGAC TAGAGACTAC TTATTCTGAA TTCATACTGT CTTCACTGTC    3600

AGTTATTTTT GAGTCAGGTT CCAACTCTCC CAGTTTGACC CTGGGGCACT TCTTTCTAGC    3660

TAGAACACAC ATGTACTCAT TTTTCAGGGT TTAAATGCAG CAGTGCCCTG CAACTTGGGC    3720

CATCTTGGAC CTTACAGCAG AGTTTCTGCT TCAGGAGCTT CTCAGAAAGT GCAGCCTTGT    3780
```

```
CTGTAGTAAG  TAAACCATCC  TTCACTGGTG  GCTGTTGAGC  CTCTGAGGAA  ATCTTGCCAC      3840

GATCAAGAGG  TCAGCATCTT  ACAAGGAGAT  TATTTTCTTC  CCTAGACGTC  TCTTTCCATC      3900

AGAAAATTCT  TCTTGCTCAT  CTAAAATGGG  AGTTTCTGGA  ACATTTACCC  ATCAGATCTA      3960

GTTCTTTTTC  CCCTATAACT  ACCCAGTTTA  CATTAATTTA  AGGTACTCTT  TCAAAAGTCA      4020

AATGTAAAGA  TCTAGGGTTA  ACATTAGAGC  ACACATTTGT  ATTCTATTGG  GACTCTCCTC      4080

CCTAGTTCCT  TTTCTTCTCT  TGTTCTTCAG  TTCACTAGTC  CTCTGGGCAG  TCTTTTAATT      4140

GTGTCTTTTG  ACAGTGCAGC  CATCAACTAT  GATCTGGACA  AAGCAGGTAA  GAACCAAAGT      4200

CATCTTTAAT  CCATTTTGTT  TTGGTTTGGC  TTTCAGTCCT  AGATACCTCT  CTTCCATTGG      4260

TATTGCTTCT  TCAATCATCA  GAATGAAAGA  AACACTTATT  AGTATCCTTC  TCTTGTCCAA      4320

TTTCCAGCTT  TTAGTAAATA  AAACATTTCT  TCTTTAATCA  GACTGACTAA  AATTTAAAGA      4380

ATGAGAAACC  CAACAGCACA  ATGCATCTTG  TGACATACAT  ATTAAGGGAG  TTTTTAATCA      4440

CAGTCAAGTC  TCCAGGACAG  AAAGAATAAT  AATTTGGTCA  TTCTGGCTAG  TAGCATAGAA      4500

TGGTTTCTGG  TGCAGTGCCG  GAGGCATTCT  AGTATTCAGG  CAGGAAGTTT  CCCACTGACC      4560

ATAAGCTACT  TATTTGCCTT  TCTCTGAAGA  ACCTGAGAAG  CTTTGTTCCT  CCAAGGTCAA      4620

GTCTCCTAAC  ACTAGGAGGA  AAGGTTGCTC  TCGATCTTCA  CGGCCCACTC  AGACTTGCCA      4680

GGTACATGGT  CTGTTTTCTG  TATAGTCTAC  AGTCTAGACA  TTTAGTGTTT  ATGTCTGCCT      4740

TCCAAGGTTA  TTTTGACCAG  GGGTATCAAA  AGTCACTGAG  CACCATTGTT  TCCCATGTTA      4800

TGCCAGGCTT  GGTTGGCTAG  GGGCTATGTT  AAAGGCTTTA  TATGAGCCCT  TCCAGGCAGG      4860

AGCTTGAAGG  TCCATTTAAT  TTAAGGTCAC  AGGTGGCTGT  TGCATCACAG  AACTGGAAAT      4920

TGAGGGCTTA  ACTTCACAGA  AAAATCCATG  GTTGAGACAA  TTTAATGTCT  AATGACATGG      4980

AGAGTAAAAG  TGATCCTCTG  GGCAAACTAA  GCATGGAGTC  ATAATATTAT  GAAGAGCATT      5040

AGAAGCAACA  TTAGGAACAG  AATATGCCAG  CTGGAAATAC  AGGCTTGATC  AACTTCTGAA      5100

AGATCCATTT  CTATCTAAAG  TAAATACATG  AGAAGGTCAT  TATACCCAAG  TGAATGTGAA      5160

GGCCTACACT  GCCTTCCGTC  TGAATCTTTT  GCAAACCTAA  CATCTAAAGC  TTGACTGACA      5220

AGTTCATATT  CATCTACTTC  TATTTTCCTA  GGACATTGAA  TTTAGGTTAA  GCAGTAATAG      5280

ATCAAGAAGA  GGGAGTCCAG  GATTTGCTTA  CCAGTATAAA  TATATGTGGT  ATCTAAACTG      5340

GGAAGCAGTT  CAATTAGTTT  TAGGTGAGGC  ATTCATGTTG  TATGTGTGCA  TATGTATGAG      5400

TCAAATCCAG  TTAGGTCTTA  GAATCAGAAT  ACATAGTGTA  GACACTATTA  TGTAGCTCAT      5460

ATATAGCAGA  ATATGTCATA  TTTGGCTAAC  CCAAGTGGGC  ATGCTATGAG  GACTAACAGG      5520

GCAGATAATG  TTGTGACATC  ACAAACTATA  CATTGCATTT  CTGTGGGTCT  CAGCTAGCCA      5580

TCAGCTATTA  GGTTTTAATA  CCATGAAAGT  AAAGGTTATG  GGGTTGTTTC  TTGTTTGTTT      5640
```

```
GATTTGTTTT GAACTTTTCT CTCACTACTG CTTCTCTGGG AAAAACAGAT TTTGGGAGAA      5700

TGAATCAGTT AACTGAAAGT AGGTTTATTA ATACTCTTGT ACCAGCTTAT CTCTCAATTA      5760

GCAATGGGTC TTACAGTAAA TATCTGAGTT TCTACTTTGC TGCACTGAAA TACTGTGCCT      5820

TTATTATAGA ACTTTGAAAG GGTTTCCCAT GAGCTGGATG GTGGGAGCCA CACTATATGT      5880

GCCATGCTGG ACATCAGGCC AGTACTGCAA TGCAGTGGTG TATTTGATGT CTTTCAGGTG      5940

ACTTCATGGT CATGACATTC TTCTTCAATG TAAGCAGGAG GTTTGGCTTC ATTGTCTTTC      6000

AAAACTATAT CCCTTCATCT GTTACCACAA TGCTTTCCTG GGTGTCCTTC TGGATCAAGA      6060

TAGAAGCTGC TGCTGCCAGG GCCTCTGTAG GTAAGAAAGT ATGGTGATCT TAAATGTGAT      6120

CATATGTCAT AGATTAACAT TGCCTTCTCC TGTTGTCCTG CAGCACTGTC AGGAGCTTAG      6180

CCCACCTTCA CTCCTTGCTC TTATAGGGGT CAGTTCTGTG CTCACCATGG CCACACTGGG      6240

TACCTTTTCT CGTAAGAATT TCCCTCGTGT CTCCTATCTC ACAGCTTTGG ACTTCTATAT      6300

TGCAATTTGT TTCGTCTTGT GCTTCTGTAC TCTACTAGAG TTCACTGTGC TCAACTTCCT      6360

GACCTACAAT AATATTGAAC GACAGGCTTC TCCAAAGTTC TACCAAGTAA GAACCACGTG      6420

TTGGTATGGG AAGCAAAACT TTAGACTGAA GATAGGAGGG TTATGGCACT TGTGTGACTA      6480

CCTTTCCAAG TGATGAGAAT TTCCCTGGGG TTATAGAGGA AATAAACTTG CTGTGATATC      6540

CAGGAAGGCC TAGAACTGAT GTGGAGAGTG TAGTTTCTAG GAAATTCTAG GCAACTTTTC      6600

TTCTTGACTC ACTTTCCCTT GACCATTTTG TTATCTTTAT TTTTCTATTT CAGTTTCCAA      6660

CCAATAGCCG TGCTAATGCA CGTACTCGTG CTCGTGCCCG CACTCGTGCT CGTGCTCGTG      6720

CCCGTGCTCG TCAGCAGCAG GAAGTGTTTG TGTGTGAGAT TGTTACCTAT GAGGAAAATG      6780

CTGAAGAGGG TTACCAGTGG TCTCCAAGAT CAAGAAGACC TCAGTGTCCC TGGAGGCGAT      6840

GTGGCCGAAG CTATGTGTGC TTCAGGGTTC TCAGGAAGTA TTTCTGCATG GTTCCTGGTT      6900

GTGAGGGCAA CAGCTGGCAG CGGGGCCGCA TCTGCATCCA CGTTTATCGC CTGGATAACT      6960

ACTCGCGGGT GCTTTTCCCC ATTACATTCT TCTTCTTTAA TGTGGTCTAC TGGGTGATTT      7020

GCCTTAACCT GTAGGCTCCA GCTGGTAGCT CATGGGGCAA TCACCTCAGT TCCCCAGGAG      7080

GTCCTAAGCC CCTTTGTCAA GGGAGTTGGG AGACAATAGC AACAGCAGCA AGCCTGAGGA      7140

GAGTTTGTCT TTGCTGCTCC TCCTTTTGTT GGCCCCTTCA CTCAATCTTT TAAACAGCAC      7200

TTCTCAAGTA ACAGCCCACC TCTCTATTCT TCAAAGAGTA TCCATGATGT TCAGTGTGCC      7260

CACAATTGAG CAGATAAGCT ATCTCTTGGT TGTGTTTGCA GTTACTGTTT TTCCCCAGGA      7320

GGTCATGGTA CCATTTGTGG TGCTAAGGGG CCCAGCTCAA GCTCAGCCCT AGAATGCACA      7380

GAGCAACAAC AGTATAGCAG AAGGAAGGTC CTCATCCTTA AGTCTTTAAG TCAGGCTATT      7440

TGATTTTTTC TCCCTGCTCT CTCTTCCTCT ACACACACA                            7479
```

(2) INFORMATION FOR SEQ ID NO:7:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 10 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CGCGACCTCC   10

(2) INFORMATION FOR SEQ ID NO:8:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 10 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TCCAGTCGAG   10

(2) INFORMATION FOR SEQ ID NO:9:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 10 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GTGAGTCTCC   10

(2) INFORMATION FOR SEQ ID NO:10:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 10 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

  (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AATGTTATAG   10

(2) INFORMATION FOR SEQ ID NO:11:

     (i) SEQUENCE CHARACTERISTICS:
         (A) LENGTH: 10 base pairs
         (B) TYPE: nucleic acid
         (C) STRANDEDNESS: double
         (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GGTCGAGGGA   10

(2) INFORMATION FOR SEQ ID NO:12:

     (i) SEQUENCE CHARACTERISTICS:
         (A) LENGTH: 10 base pairs
         (B) TYPE: nucleic acid
         (C) STRANDEDNESS: double
         (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GGCATTGGAG   10

(2) INFORMATION FOR SEQ ID NO:13:

     (i) SEQUENCE CHARACTERISTICS:
         (A) LENGTH: 10 base pairs
         (B) TYPE: nucleic acid
         (C) STRANDEDNESS: double
         (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GTGAGGAGCA   10

(2) INFORMATION FOR SEQ ID NO:14:

     (i) SEQUENCE CHARACTERISTICS:
         (A) LENGTH: 10 base pairs
         (B) TYPE: nucleic acid
         (C) STRANDEDNESS: double

        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TCCCTTATAG   10

(2) INFORMATION FOR SEQ ID NO:15:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 10 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: cDNA

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AGAAGCCCAC   10

(2) INFORMATION FOR SEQ ID NO:16:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 10 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: cDNA

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CCTAGACATG   10

(2) INFORMATION FOR SEQ ID NO:17:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 10 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: cDNA

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GTGAGTACTA   10

(2) INFORMATION FOR SEQ ID NO:18:

 (i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 10 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: double
  (D) TOPOLOGY: linear

 (ii) MOLECULE TYPE: cDNA

 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TTCCCACCAG   10

(2) INFORMATION FOR SEQ ID NO:19:

 (i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 10 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: double
  (D) TOPOLOGY: linear

 (ii) MOLECULE TYPE: cDNA

 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GAATACACCA   10

(2) INFORMATION FOR SEQ ID NO:20:

 (i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 10 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: double
  (D) TOPOLOGY: linear

 (ii) MOLECULE TYPE: cDNA

 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

ACACAATTAG   10

(2) INFORMATION FOR SEQ ID NO:21:

 (i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 10 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: double
  (D) TOPOLOGY: linear

 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GTATGTCAAG   10

(2) INFORMATION FOR SEQ ID NO:22:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 10 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: cDNA

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TCCCTTCCAG   .   10

(2) INFORMATION FOR SEQ ID NO:23:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 10 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: cDNA

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GATGACCATT   10

(2) INFORMATION FOR SEQ ID NO:24:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 10 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: cDNA

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TTCTCTAGCT   10

(2) INFORMATION FOR SEQ ID NO:25:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 10 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double

          (D) TOPOLOGY: linear

     (ii) MOLECULE TYPE: cDNA

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GTGAGTACCT   10

(2) INFORMATION FOR SEQ ID NO:26:

          (i) SEQUENCE CHARACTERISTICS:
               (A) LENGTH: 10 base pairs
               (B) TYPE: nucleic acid
               (C) STRANDEDNESS: double
               (D) TOPOLOGY: linear

          (ii) MOLECULE TYPE: cDNA

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TCCTTTCCAG   10

(2) INFORMATION FOR SEQ ID NO:27:

          (i) SEQUENCE CHARACTERISTICS:
               (A) LENGTH: 10 base pairs
               (B) TYPE: nucleic acid
               (C) STRANDEDNESS: double
               (D) TOPOLOGY: linear

          (ii) MOLECULE TYPE: cDNA

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TTTCCTATCC   10

(2) INFORMATION FOR SEQ ID NO:28:

          (i) SEQUENCE CHARACTERISTICS:
               (A) LENGTH: 10 base pairs
               (B) TYPE: nucleic acid
               (C) STRANDEDNESS: double
               (D) TOPOLOGY: linear

          (ii) MOLECULE TYPE: cDNA

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

ACCCCAGTTG   10

(2) INFORMATION FOR SEQ ID NO:29:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 10 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: cDNA

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GTAAGCGTGC   10

(2) INFORMATION FOR SEQ ID NO:30:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 10 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: cDNA

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GTCTTTTCAG   10

(2) INFORMATION FOR SEQ ID NO:31:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 10 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: cDNA

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GTGACTTCAT   10

(2) INFORMATION FOR SEQ ID NO:32:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 10 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

ACCTCTCTAG   10

(2) INFORMATION FOR SEQ ID NO:33:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 10 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GTAAGAGGAG   10

(2) INFORMATION FOR SEQ ID NO:34:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 10 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GCTCTTGCAG   10

(2) INFORMATION FOR SEQ ID NO:35:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 10 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GGATCACCTC   10

(2) INFORMATION FOR SEQ ID NO:36:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 10 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double

        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

ACTCCGCCAT   .   10

(2) INFORMATION FOR SEQ ID NO:37:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 10 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GTATGAGCTG   10

(2) INFORMATION FOR SEQ ID NO:38:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 10 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

TCCATTTTAG   10

(2) INFORMATION FOR SEQ ID NO:39:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 10 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CCTCGTATCA   10

(2) INFORMATION FOR SEQ ID NO:40:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 10 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

ATATATCAGC   10

(2) INFORMATION FOR SEQ ID NO:41:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 3154 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: both
        (D) TOPOLOGY: both

    (ii) MOLECULE TYPE: cDNA

    (ix) FEATURE:
        (A) NAME/KEY: CDS
        (B) LOCATION: 41..1558

    (ix) FEATURE:
        (A) NAME/KEY: sig_peptide
        (B) LOCATION: 41..94

    (ix) FEATURE:
        (A) NAME/KEY: mat_peptide
        (B) LOCATION: 95..1558

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

```
CGCGACCTCC GCGCAGGTGG TCGCGCCGGT CTCCGCGGAA ATG TTG TCC AAA GTT        55
                                             Met Leu Ser Lys Val
                                             -18          -15

CTT CCA GTC CTC CTA GGC ATC TTA TTG ATC CTC CAG TCG AGG GTC GAG        103
Leu Pro Val Leu Leu Gly Ile Leu Leu Ile Leu Gln Ser Arg Val Glu
        -10              -5                  1

GGA CCT CAG ACT GAA TCA AAG AAT GAA GCC TCT TCC CGT GAT GTT GTC        151
Gly Pro Gln Thr Glu Ser Lys Asn Glu Ala Ser Ser Arg Asp Val Val
    5               10                  15

TAT GGC CCC CAG CCC CAG CCT CTG GAA AAT CAG CTC CTC TCT GAG GAA        199
Tyr Gly Pro Gln Pro Gln Pro Leu Glu Asn Gln Leu Leu Ser Glu Glu
20                  25                  30                  35

ACA AAG TCA ACT GAG ACT GAG ACT GGG AGC AGA GTT GGC AAA CTG CCA        247
Thr Lys Ser Thr Glu Thr Glu Thr Gly Ser Arg Val Gly Lys Leu Pro
                40                  45                  50
```

```
GAA GCC TCT CGC ATC CTG AAC ACT ATC CTG AGT AAT TAT GAC CAC AAA        295
Glu Ala Ser Arg Ile Leu Asn Thr Ile Leu Ser Asn Tyr Asp His Lys
            55                  60                  65

CTG CGC CCT GGC ATT GGA GAG AAG CCC ACT GTG GTC ACT GTT GAG ATC        343
Leu Arg Pro Gly Ile Gly Glu Lys Pro Thr Val Val Thr Val Glu Ile
            70                  75                  80

GCC GTC AAC AGC CTT GGT CCT CTC TCT ATC CTA GAC ATG GAA TAC ACC        391
Ala Val Asn Ser Leu Gly Pro Leu Ser Ile Leu Asp Met Glu Tyr Thr
            85                  90                  95

ATT GAC ATC ATC TTC TCC CAG ACC TGG TAC GAC GAA CGC CTC TGT TAC        439
Ile Asp Ile Ile Phe Ser Gln Thr Trp Tyr Asp Glu Arg Leu Cys Tyr
100                 105                 110                 115

AAC GAC ACC TTT GAG TCT CTT GTT CTG AAT GGC AAT GTG GTG AGC CAG        487
Asn Asp Thr Phe Glu Ser Leu Val Leu Asn Gly Asn Val Val Ser Gln
                120                 125                 130

CTA TGG ATC CCG GAC ACC TTT TTT AGG AAT TCT AAG AGG ACC CAC GAG        535
Leu Trp Ile Pro Asp Thr Phe Phe Arg Asn Ser Lys Arg Thr His Glu
                135                 140                 145

CAT GAG ATC ACC ATG CCC AAC CAG ATG GTC CGC ATC TAC AAG GAT GGC        583
His Glu Ile Thr Met Pro Asn Gln Met Val Arg Ile Tyr Lys Asp Gly
            150                 155                 160

AAG GTG TTG TAC ACA ATT AGG ATG ACC ATT GAT GCC GGA TGC TCA CTC        631
Lys Val Leu Tyr Thr Ile Arg Met Thr Ile Asp Ala Gly Cys Ser Leu
            165                 170                 175

CAC ATG CTC AGA TTT CCA ATG GAT TCT CAC TCT TGC CCT CTA TCT TTC        679
His Met Leu Arg Phe Pro Met Asp Ser His Ser Cys Pro Leu Ser Phe
180                 185                 190                 195

TCT AGC TTT TCC TAT CCT GAG AAT GAG ATG ATC TAC AAG TGG GAA AAT        727
Ser Ser Phe Ser Tyr Pro Glu Asn Glu Met Ile Tyr Lys Trp Glu Asn
                200                 205                 210

TTC AAG CTT GAA ATC AAT GAG AAG AAC TCC TGG AAG CTC TTC CAG TTT        775
Phe Lys Leu Glu Ile Asn Glu Lys Asn Ser Trp Lys Leu Phe Gln Phe
            215                 220                 225

GAT TTT ACA GGA GTG AGC AAC AAA ACT GAA ATA ATC ACA ACC CCA GTT        823
Asp Phe Thr Gly Val Ser Asn Lys Thr Glu Ile Ile Thr Thr Pro Val
            230                 235                 240

GGT GAC TTC ATG GTC ATG ACG ATT TTC TTC AAT GTG AGC AGG CGG TTT        871
Gly Asp Phe Met Val Met Thr Ile Phe Phe Asn Val Ser Arg Arg Phe
            245                 250                 255

GGC TAT GTT GCC TTT CAA AAC TAT GTC CCT TCT TCC GTG ACC ACG ATG        919
Gly Tyr Val Ala Phe Gln Asn Tyr Val Pro Ser Ser Val Thr Thr Met
260                 265                 270                 275

CTC TCC TGG GTT TCC TTT TGG ATC AAG ACA GAG TCT GCT CCA GCC CGG        967
Leu Ser Trp Val Ser Phe Trp Ile Lys Thr Glu Ser Ala Pro Ala Arg
                280                 285                 290

ACC TCT CTA GGG ATC ACC TCT GTT CTG ACC ATG ACC ACG TTG GGC ACC       1015
Thr Ser Leu Gly Ile Thr Ser Val Leu Thr Met Thr Thr Leu Gly Thr
```

```
                  295                        300                        305

TTT TCT CGT AAG AAT TTC CCG CGT GTC TCC TAT ATC ACA GCC TTG GAT        1063
Phe Ser Arg Lys Asn Phe Pro Arg Val Ser Tyr Ile Thr Ala Leu Asp
        310                        315                        320

TTC TAT ATC GCC ATC TGC TTC GTC TTC TGC TTC TGC GCT CTG TTG GAG        1111
Phe Tyr Ile Ala Ile Cys Phe Val Phe Cys Phe Cys Ala Leu Leu Glu
        325                        330                        335

TTT GCT GTG CTC AAC TTC CTG ATC TAC AAC CAG ACA AAA GCC CAT GCT        1159
Phe Ala Val Leu Asn Phe Leu Ile Tyr Asn Gln Thr Lys Ala His Ala
340                        345                        350                355

TCT CCT AAA CTC CGC CAT CCT CGT ATC AAT AGC CGT GCC CAT GCC CGT        1207
Ser Pro Lys Leu Arg His Pro Arg Ile Asn Ser Arg Ala His Ala Arg
                360                        365                        370

ACC CGT GCA CGT TCC CGA GCC TGT GCC CGC CAA CAT CAG GAA GCT TTT        1255
Thr Arg Ala Arg Ser Arg Ala Cys Ala Arg Gln His Gln Glu Ala Phe
                375                        380                        385

GTG TGC CAG ATT GTC ACC ACT GAG GGA AGT GAT GGA GAG GAG CGC CCG        1303
Val Cys Gln Ile Val Thr Thr Glu Gly Ser Asp Gly Glu Glu Arg Pro
                390                        395                        400

TCT TGC TCA GCC CAG CAG CCC CCT AGC CCA GGT AGC CCT GAG GGT CCC        1351
Ser Cys Ser Ala Gln Gln Pro Pro Ser Pro Gly Ser Pro Glu Gly Pro
                405                        410                        415

CGC AGC CTC TGC TCC AAG CTG GCC TGC TGT GAG TGG TGC AAG CGT TTT        1399
Arg Ser Leu Cys Ser Lys Leu Ala Cys Cys Glu Trp Cys Lys Arg Phe
420                        425                        430                435

AAG AAG TAC TTC TGC ATG GTC CCC GAT TGT GAG GGC AGT ACC TGG CAG        1447
Lys Lys Tyr Phe Cys Met Val Pro Asp Cys Glu Gly Ser Thr Trp Gln
                440                        445                        450

CAG GGC CGC CTC TGC ATC CAT GTC TAC CGC CTG GAT AAC TAC TCG AGA        1495
Gln Gly Arg Leu Cys Ile His Val Tyr Arg Leu Asp Asn Tyr Ser Arg
                455                        460                        465

GTT GTT TTC CCA GTG ACT TTC TTC TTC TTC AAT GTG CTC TAC TGG CTT        1543
Val Val Phe Pro Val Thr Phe Phe Phe Phe Asn Val Leu Tyr Trp Leu
                470                        475                        480

GTT TGC CTT AAC TTG TAG GTACCAGCTG GTACCCTGTG GGGCAACCTC             1591
Val Cys Leu Asn Leu  *
                485

TCCAGTTCCC CAGGAGGTCC AAGCCCCTTG CCAAGGGAGT TGGGGGAAAG CAGCAGCAGC     1651

AGCAGGAGCG ACTAGAGTTT TTCCTGCCCC ATTCCCCAAA CAGAAGCTTG CAGAGGGTTT     1711

GTCTTTGCTG CCCCTCTCCC CTACCTGGCC CATTCACTGA GTCTTCTCAG CAGACCATTT     1771

CAAATTATTA ATAAATGGGC CACCTCCCTC TTCTTCAAGG AGCATCCGTG ATGCTCAGTG     1831

TTCAAAACCA CAGCCACTTA GTGATCAGCT CCCTAAAACC ATGCCTAAGT ACAGGCGGAT     1891

TAGCTATCTT CCAACAATGC TGACCACCAG ACAATTACTG CATTTTTCCA GAAGCCCACT     1951
```

```
ATTGCCTTTG TAGTGCTTTC GGCCCAGTTC TGGCCTCAGC CTCAAAGTGC ACCGACTAGT    2011

TGCTTGCCTA TACCTGGCAC CTCATTAAGA TGCTGGGCAG CAGTATAACA GGAGGAAGAG    2071

ATCCCTCTCC TTTGGTCAGA TTATTATGTT CTCAGTTCTC TCTCCCTGCT ACCCCTTTCT    2131

CTGCAGATAG ATAGACACTG GCATTATCCC TTTAGGAAGA GGGGGGGGCA GCAAGAGAGC    2191

CTATTTGGGA CAGCATTCCT CTCTCTCTGC TGCTGTGACA TCTCCCTCTC CTTGCTGGCT    2251

CCATCTTTCG TCTGCACTAC CAATTCAATG CCCTTCATCC AATGGGTATC TATTTTTGTG    2311

TGTGATTATA GTAACTACTC CCTGCTTTAT ATGCCACCCT CTTCCTTCTC TTTGACCCCT    2371

GTGACTCTTT CTGTAACTTT CCCAGTGACT TCCCCTAGCC CTGACCCAGG CACTAGGCCT    2431

TGGTGACTTC CTGGGGCCAA GAAACTAAGG AAACTCGGCT TTGCAACAGG CATTACTCGC    2491

CATTGATTGG TGCCCACCCA GGGCACACTG TCGGAGTTCT ATCACTTGCT TGACCCCTGG    2551

ACCCATAAAC CAGTCCACTG TTATACCCGG GGCACTCTAA CCATCACAAT CAATCAATCA    2611

AATTCCCTTA AATTTGTATG GCACTGGAAC TTTGGCAAAG CACTTTTGAC AAGTTGTGTC    2671

TGATTGGAGC TTCATGATAG CCTTGTGACA TCTTTAGGGC AGGATTCTTA TCCCCATTTT    2731

GCAGATGAAA ACCCTGAGTC ACAGATTTCT GTGGGACTGT GGATCTCACT GGAAGCTATC    2791

CAAGAGCCCA CTGTCACCTT CTAGACCACA TGATAGGGCT AGACAGCTCA GTTCACCATG    2851

ATTCTCTTCT GTCACCTCTG CTGGCACACC AGTGGCAAGG CCCAGAATGG CGACCTCTCT    2911

TTAGCTCAAT TTCTGGGCCT GAGGTGCTCA GACTGCCCCC AAGATCAAAT CTCTCCTGGC    2971

TGTAGTAACC CAGTGGAATG AATTTGGACA TGCCCCAATG CTTCTATATG CTAAGTGAAA    3031

TCTGTGTCTG TAATTTGTTG GGGGGTGGAT AGGGTGGGGT CTCCATCTAC TTTTTGTCAC    3091

CATCATCTGA AATGGGGAAA TATGTAAATA AATATATCAG CAAAGCAAAA AGAAAAAAAA    3151

AAA                                                                  3154
```

(2) INFORMATION FOR SEQ ID NO:42:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 506 amino acids
            (B) TYPE: amino acid
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: protein

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

```
Met Leu Ser Lys Val Leu Pro Val Leu Leu Gly Ile Leu Leu Ile Leu
-18          -15              -10                  -5

Gln Ser Arg Val Glu Gly Pro Gln Thr Glu Ser Lys Asn Glu Ala Ser
         1               5                  10

Ser Arg Asp Val Val Tyr Gly Pro Gln Pro Gln Pro Leu Glu Asn Gln
 15                  20              25                  30
```

Leu Leu Ser Glu Glu Thr Lys Ser Thr Glu Thr Glu Thr Gly Ser Arg
            35   40   45

Val Gly Lys Leu Pro Glu Ala Ser Arg Ile Leu Asn Thr Ile Leu Ser
            50   55   60

Asn Tyr Asp His Lys Leu Arg Pro Gly Ile Gly Glu Lys Pro Thr Val
        65   70   75

Val Thr Val Glu Ile Ala Val Asn Ser Leu Gly Pro Leu Ser Ile Leu
        80   85   90

Asp Met Glu Tyr Thr Ile Asp Ile Ile Phe Ser Gln Thr Trp Tyr Asp
95   100   105   110

Glu Arg Leu Cys Tyr Asn Asp Thr Phe Glu Ser Leu Val Leu Asn Gly
                115   120   125

Asn Val Val Ser Gln Leu Trp Ile Pro Asp Thr Phe Phe Arg Asn Ser
            130   135   140

Lys Arg Thr His Glu His Glu Ile Thr Met Pro Asn Gln Met Val Arg
        145   150   155

Ile Tyr Lys Asp Gly Lys Val Leu Tyr Thr Ile Arg Met Thr Ile Asp
    160   165   170

Ala Gly Cys Ser Leu His Met Leu Arg Phe Pro Met Asp Ser His Ser
175   180   185   190

Cys Pro Leu Ser Phe Ser Ser Phe Ser Tyr Pro Glu Asn Glu Met Ile
            195   200   205

Tyr Lys Trp Glu Asn Phe Lys Leu Glu Ile Asn Glu Lys Asn Ser Trp
        210   215   220

Lys Leu Phe Gln Phe Asp Phe Thr Gly Val Ser Asn Lys Thr Glu Ile
        225   230   235

Ile Thr Thr Pro Val Gly Asp Phe Met Val Met Thr Ile Phe Phe Asn
    240   245   250

Val Ser Arg Arg Phe Gly Tyr Val Ala Phe Gln Asn Tyr Val Pro Ser
255   260   265   270

Ser Val Thr Thr Met Leu Ser Trp Val Ser Phe Trp Ile Lys Thr Glu
            275   280   285

Ser Ala Pro Ala Arg Thr Ser Leu Gly Ile Thr Ser Val Leu Thr Met
            290   295   300

Thr Thr Leu Gly Thr Phe Ser Arg Lys Asn Phe Pro Arg Val Ser Tyr
        305   310   315

Ile Thr Ala Leu Asp Phe Tyr Ile Ala Ile Cys Phe Val Phe Cys Phe
    320   325   330

Cys Ala Leu Leu Glu Phe Ala Val Leu Asn Phe Leu Ile Tyr Asn Gln
335   340   345   350

Thr Lys Ala His Ala Ser Pro Lys Leu Arg His Pro Arg Ile Asn Ser
            355   360   365

```
Arg Ala His Ala Arg Thr Arg Ala Arg Ser Arg Ala Cys Ala Arg Gln
            370             375             380

His Gln Glu Ala Phe Val Cys Gln Ile Val Thr Thr Glu Gly Ser Asp
            385             390             395

Gly Glu Glu Arg Pro Ser Cys Ser Ala Gln Gln Pro Pro Ser Pro Gly
    400             405             410

Ser Pro Glu Gly Pro Arg Ser Leu Cys Ser Lys Leu Ala Cys Cys Glu
415             420             425             430

Trp Cys Lys Arg Phe Lys Lys Tyr Phe Cys Met Val Pro Asp Cys Glu
            435             440             445

Gly Ser Thr Trp Gln Gln Gly Arg Leu Cys Ile His Val Tyr Arg Leu
            450             455             460

Asp Asn Tyr Ser Arg Val Val Phe Pro Val Thr Phe Phe Phe Phe Asn
            465             470             475

Val Leu Tyr Trp Leu Val Cys Leu Asn Leu
480             485
```

(2) INFORMATION FOR SEQ ID NO:43:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 303 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: both
        (D) TOPOLOGY: both

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

```
TAGGTACCAG CTGGTACCCT GTGGGGCAAC CTCTCCAGTT CCCCAGGAGG TCCAAGCCCC    60

TTGCCAAGGG AGTTGGGGGA AAGCAGCAGC AGCAGCAGGA GCGACTAGAG TTTTTCCTGC   120

CCCATTCCCC AAACAGAAGC TTGCAGAGGG TTTGTTTTTG CTGCCCCTCT CCCCTACCTG   180

GCCCATTCAC TGAGTCTTCT CAGCAGACCA TTTCAAATTA TTAATAAATG GGCCACCTCC   240

CTCTTCTTCA AGGAGCATCC GTGATGCTCA GTGTTCAAAA CCACAGCCAC TTAGTGATCA   300

GCT                                                                303
```

(2) INFORMATION FOR SEQ ID NO:44:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 206 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: both
        (D) TOPOLOGY: both

    (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

```
GCTTTGCTGA TATATTTATT TACATATTTC CCCATTTCAG ATGATGGTGA CAAAAAGTAG        60

ATGGAGACCC CACCCTATCC ACCCCCCAAC AAATTACAGA CACAGATTTC ACTTAGCATA       120

TAGAAGCATT GGGGCATGTC CAAATTCATT CCACTGGGTT ACTACAGCCA GGAGAGATTT       180

GATCTTGGGG GCAGTCTGGA GCACCT                                           206
```

(2) INFORMATION FOR SEQ ID NO:45:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 3974 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: both
        (D) TOPOLOGY: both

    (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

```
GGTACCTAAG TGAGTAGGGC GTCCGATCGA CGGACGCCTT TTTTTTGAAT TCGTAATCAT         60

GGTCATAGCT GTTTCCTGTG TGAAATTGTT ATCCGCTCAC AATTCCACAC AACATACGAG        120

CCGGAAGCAT AAAGTGTAAA GCCTGGGGTG CCTAATGAGT GAGCTAACTC ACATTAATTG        180

CGTTGCGCTC ACTGCCCGCT TTCCAGTCGG GAAACCTGTC GTGCCAGCTG CATTAATGAA        240

TCGGCCAACG CGCGGGGAGA GGCGGTTTGC GTATTGGGCG CTCTTCCGCT TCCTCGCTCA        300

CTGACTCGCT GCGCTCGGTC GTTCGGCTGC GGCGAGCGGT ATCAGCTCAC TCAAAGGCGG        360

TAATACGGTT ATCCACAGAA TCAGGGGATA ACGCAGGAAA GAACATGTGA GCAAAAGGCC        420

AGCAAAAGGC CAGGAACCGT AAAAAGGCCG CGTTGCTGGC GTTTTTCCAT AGGCTCCGCC        480

CCCCTGACGA GCATCACAAA AATCGACGCT CAAGTCAGAG GTGGCGAAAC CGACAGGAC         540

TATAAAGATA CCAGGCGTTT CCCCCTGGAA GCTCCCTCGT GCGCTCTCCT GTTCCGACCC        600

TGCCGCTTAC CGGATACCTG TCCGCCTTTC TCCCTTCGGG AAGCGTGGCG CTTTCTCATA        660

GCTCACGCTG TAGGTATCTC AGTTCGGTGT AGGTCGTTCG CTCCAAGCTG GGCTGTGTGC        720

ACGAACCCCC CGTTCAGCCC GACCGCTGCG CCTTATCCGG TAACTATCGT CTTGAGTCCA        780

ACCCGGTAAG ACACGACTTA TCGCCACTGG CAGCAGCCAC TGGTAACAGG ATTAGCAGAG        840

CGAGGTATGT AGGCGGTGCT ACAGAGTTCT TGAAGTGGTG GCCTAACTAC GGCTACACTA        900

GAAGAACAGT ATTTGGTATC TGCGCTCTGC TGAAGCCAGT TACCTTCGGA AAAAGAGTTG        960

GTAGCTCTTG ATCCGGCAAA CAAACCACCG CTGGTAGCGG TGGTTTTTTT GTTTGCAAGC       1020

AGCAGATTAC GCGCAGAAAA AAAGGATCTC AAGAAGATCC TTTGATCTTT TCTACGGGGT       1080
```

```
CTGACGCTCA GTGGAACGAA AACTCACGTT AAGGGATTTT GGTCATGAGA TTATCGTCGA    1140

CAATTCGCGC GCGAAGGCGA AGCGGCATGC ATTTACGTTG ACACCATCGA ATGGTGCAAA    1200

ACCTTTCGCG GTATGGCATG ATAGCGCCCG GAAGAGAGTC AATTCAGGGT GGTGAATGTG    1260

AAACCAGTAA CGTTATACGA TGTCGCAGAG TATGCCGGTG TCTCTTATCA GACCGTTTCC    1320

CGCGTGGTGA ACCAGGCCAG CCACGTTTCT GCGAAAACGC GGGAAAAAGT GGAAGCGGCG    1380

ATGGCGGAGC TGAATTACAT TCCCAACCGC GTGGCACAAC AACTGGCGGG CAAACAGTCG    1440

TTGCTGATTG GCGTTGCCAC CTCCAGTCTG GCCCTGCACG CGCCGTCGCA AATTGTCGCG    1500

GCGATTAAAT CTCGCGCCGA TCAACTGGGT GCCAGCGTGG TGGTGTCGAT GGTAGAACGA    1560

AGCGGCGTCG AAGCCTGTAA AGCGGCGGTG CACAATCTTC TCGCGCAACG CGTCAGTGGG    1620

CTGATCATTA ACTATCCGCT GGATGACCAG GATGCCATTG CTGTGGAAGC TGCCTGCACT    1680

AATGTTCCGG CGTTATTTCT TGATGTCTCT GACCAGACAC CCATCAACAG TATTATTTTC    1740

TCCCATGAAG ACGGTACGCG ACTGGGCGTG GAGCATCTGG TCGCATTGGG TCACCAGCAA    1800

ATCGCGCTGT TAGCGGGCCC ATTAAGTTCT GTCTCGGCGC GTCTGCGTCT GGCTGGCTGG    1860

CATAAATATC TCACTCGCAA TCAAATTCAG CCGATAGCGG AACGGGAAGG CGACTGGAGT    1920

GCCATGTCCG GTTTTCAACA AACCATGCAA ATGCTGAATG AGGGCATCGT TCCCACTGCG    1980

ATGCTGGTTG CCAACGATCA GATGGCGCTG GGCGCAATGC GCGCCATTAC CGAGTCCGGG    2040

CTGCGCGTTG GTGCGGATAT CTCGGTAGTG GGATACGACG ATACCGAAGA CAGCTCATGT    2100

TATATCCCGC CGTTAACCAC CATCAAACAG GATTTTCGCC TGCTGGGGCA AACCAGCGTG    2160

GACCGCTTGC TGCAACTCTC TCAGGGCCAG GCGGTGAAGG GCAATCAGCT GTTGCCCGTC    2220

TCACTGGTGA AAAGAAAAAC CACCCTGGCG CCCAATACGC AAACCGCCTC TCCCCGCGCG    2280

TTGGCCGATT CATTAATGCA GCTGGCACGA CAGGTTTCCC GACTGGAAAG CGGGCAGTGA    2340

GCGCAACGCA ATTAATGTAA GTTAGCGCGA ATTGTCGACC AAAGCGGCCA TCGTGCCTCC    2400

CCACTCCTGC AGTTCGGGGG CATGGATGCG CGGATAGCCG CTGCTGGTTT CCTGGATGCC    2460

GACGGATTTG CACTGCCGGT AGAACTCCGC GAGGTCGTCC AGCCTCAGGC AGCAGCTGAA    2520

CCAACTCGCG AGGGGATCGA GCCCGGGGTG GGCGAAGAAC TCCAGCATGA GATCCCCGCG    2580

CTGGAGGATC ATCCAGCCGG CGTCCCGGAA AACGATTCCG AAGCCCAACC TTTCATAGAA    2640

GGCGGCGGTG GAATCGAAAT CTCGTGATGG CAGGTTGGGC GTCGCTTGGT CGGTCATTTC    2700

GAACCCCAGA GTCCCGCTCA GAAGAACTCG TCAAGAAGGC GATAGAAGGC GATGCGCTGC    2760

GAATCGGGAG CGGCGATACC GTAAAGCACG AGGAAGCGGT CAGCCCATTC GCCGCCAAGC    2820

TCTTCAGCAA TATCACGGGT AGCCAACGCT ATGTCCTGAT AGCGGTCCGC CACACCCAGC    2880

CGGCCACAGT CGATGAATCC AGAAAAGCGG CCATTTTCCA CCATGATATT CGGCAAGCAG    2940
```

```
GCATCGCCAT GGGTCACGAC GAGATCCTCG CCGTCGGGCA TGCGCGCCTT GAGCCTGGCG    3000

AACAGTTCGG CTGGCGCGAG CCCCTGATGC TCTTCGTCCA GATCATCCTG ATCGACAAGA    3060

CCGGCTTCCA TCCGAGTACG TGCTCGCTCG ATGCGATGTT TCGCTTGGTG GTCGAATGGG    3120

CAGGTAGCCG GATCAAGCGT ATGCAGCCGC CGCATTGCAT CAGCCATGAT GGATACTTTC    3180

TCGGCAGGAG CAAGGTGAGA TGACAGGAGA TCCTGCCCCG GCACTTCGCC CAATAGCAGC    3240

CAGTCCCTTC CCGCTTCAGT GACAACGTCG AGCACAGCTG CGCAAGGAAC GCCCGTCGTG    3300

GCCAGCCACG ATAGCCGCGC TGCCTCGTCC TGCAGTTCAT TCAGGGCACC GGACAGGTCG    3360

GTCTTGACAA AAAGAACCGG GCGCCCCTGC GCTGACAGCC GGAACACGGC GGCATCAGAG    3420

CAGCCGATTG TCTGTTGTGC CCAGTCATAG CCGAATAGCC TCTCCACCCA AGCGGCCGGA    3480

GAACCTGCGT GCAATCCATC TTGTTCAATC ATGCGAAACG ATCCTCATCC TGTCTCTTGA    3540

TCAGATCTTG ATCCCCTGCG CCATCAGATC CTTGGCGGCA AGAAAGCCAT CCAGTTTACT    3600

TTGCAGGGCT TCCCAACCTT ACCAGAGGGC GCCCCAGCTG GCAATTCCGG TTCGCTTGCT    3660

GTCCATAAAA CCGCCCAGTC TAGCTATCGC CATGTAAGCC CACTGCAAGC TACCTGCTTT    3720

CTCTTTGCGC TTGCGTTTTC CCTTGTCCAG ATAGCCCAGT AGCTGACATT CATCCGGGGT    3780

CAGCACCGTT TCTGCGGACT GGCTTTCTAC GTGTTCCGCT TCCTTTAGCA GCCCTTGCGC    3840

CCTGAGTGCT TGCGGCAGCG TGAAGCTTAA AAAACTGCAA AAAATAGTTT GACTTGTGAG    3900

CGGATAACAA TTAAGATGTA CCCAATTGTG AGCGGATAAC AATTTCACAC ATTAAAGAGG    3960

AGAAATTACA TATG                                                      3974
```

(2) INFORMATION FOR SEQ ID NO:46:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 112 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: both
        (D) TOPOLOGY: both

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

```
AAGCTTAAAA AACTGCAAAA AATAGTTTGA CTTGTGAGCG GATAACAATT AAGATGTACC      60

CAATTGTGAG CGGATAACAA TTTCACACAT TAAAGAGGAG AAATTACATA TG             112
```